

REMARKS/ARGUMENTS

Favorable reconsideration of this application as presently amended and in light of the following discussion is respectfully requested.

Claims 1-4, 6-18, and 20-27 are pending in this application. Claims 5 and 19 are canceled, and Claims 1, 13-15, 26, and 27 are amended by the present amendment.

In the outstanding Office Action, Claims 15-26 were rejected under 35 U.S.C. § 101; Claims 1-9, 11-23, and 25-27 were rejected under 35 U.S.C. § 102(e) as anticipated by U.S. Patent 6,574,595 to Mitchell et al. (herein "Mitchell"); and Claims 10 and 24 were rejected under 35 U.S.C. § 103(a) as unpatentable over Mitchell in view of U.S. Publication 2002/0178004 to Chang et al. (herein "Chang").

Applicants respectfully traverse the rejection of Claims 15-26 under 35 U.S.C. § 101, with regard to amended independent Claims 15 and 26.

Claim 15 is amended to recite that a method is performed by a processor, and the method includes a receiving step of receiving, at the processor, a sequence of feature vectors corresponding to a speech segment. Such a processor is a particular apparatus, which is a member of another statutory category, and therefore amended Claim 15 includes steps that are tied to the apparatus. Similar amendments are made to Claim 26. Accordingly, Applicants respectfully submit that Claims 15 and 26, and claims depending therefrom, are tied to another statutory category and therefore are eligible for patent protection under 35 U.S.C. § 101. Therefore, it is respectfully requested the rejection under 35 U.S.C. § 101 be withdrawn.

Furthermore, Applicants respectfully traverse the rejection of Claims 1-9, 11-23, and 25-27 under 35 U.S.C. § 102(e) as anticipated by Mitchell.

Amended Claim 1 is directed to a decoder for an automatic speech recognition system for determining one or more candidate text unit concatenations according to a predetermined

criterion and which correspond to a speech segment. The decoder includes, in part, a processor arranged to receive a sequence of feature vectors corresponding to the speech segment. In addition, Claim 1 is amended to recite that a new token includes a pointer to a word history structure that includes: pointers to text unit concatenations from previous tokens of different nodes, and offsets indicating a difference in maximum likelihood scores between the text unit concatenations from the previous tokens. Independent Claims 13-15, 26, and 27 are similarly amended, and the amendments find support in the specification as originally filed at least at page 24, lines 7-11, the paragraph bridging pages 24 and 25, the paragraph bridging pages 25 and 26, and at page 26, lines 18-26. Thus, no new matter is added.

In addition, Claim 1 is amended to recite that the plurality of candidate text unit concatenations in a token are the text unit concatenations with the N-best likelihood values, for example as disclosed in original Claims 5 and 19. Thus, no new matter is added by that amendment. Moreover, Claim 1 is amended to indicate that a merging of the text unit concatenations in the new token to determine the N-best likelihood values is delayed until an end of a speech segment, and this amendment finds support in the non-limiting embodiment of Applicants' specification at page 23, lines 17-27, and the paragraph bridging pages 24 and 25. Thus, no new matter is added by the amendments to the claims.

Applicants respectfully submit that Mitchell fails to disclose each of the features of any of the independent claims. For example, it is respectfully submitted that Mitchell fails to teach or suggest that tokens are combined to generate a new token that includes a pointer to a word history structure that includes pointers to text unit concatenations from previous tokens of different nodes and offsets indicating a difference in maximum likelihood scores between the text unit concatenations from the previous tokens. In addition, it is respectfully submitted

that Mitchell fails to teach or suggest delaying, until an end of a speech segment, a merging of text unit concatenations in a new token to determine the N-best likelihood values.

According to Mitchell Figs. 2A and 2B, a new speech frame is retrieved in step 204, and various steps are performed on this frame as described by Mitchell with regard to steps 206-214.<sup>1</sup> Furthermore, according to Mitchell, in step 216 the system finds the best scoring path and “NumPhonemes” for a best path.<sup>2</sup> In particular, Mitchell indicates that “step 216 finds the best scoring path for the most likely sub-word sequence and determines the number of phonemes for sub-words in the best scoring path.”<sup>3</sup> Thus, according to Mitchell, for each frame the best likelihood scores are derived. Additionally, Mitchell indicates that after step 216 there are three decision points at steps 220, 230, and 240 of Mitchell Figure 2B to see if the utterance can be identified. However, according to Mitchell, if the decision is negative in any of these respects, then the process returns to step 204 to get a new speech frame. Therefore, it is clear that according to Mitchell, a determination of the best likelihood scores, which includes a sorting to determine the “best” likelihood scores, is performed on a per frame basis. Thus, according to Mitchell, when two tokens are merged, a sorting process is performed each time to determine the best sub-word sequences for the new token.

However, according to the amended independent claims of the present application, a merging, which may include a sorting to determine the N-best likelihood values, may not be performed each time a new token is formed from a combination of tokens, and instead, the merging is delayed until an end of a speech segment. For example, according to a non-limiting embodiment of the invention of Claim 1, a merging operation to determine the N-best likelihood values of the text concatenations is delayed until an end of the speech segment. Thus, the new token may be formed, but the known merging procedure which

---

<sup>1</sup> Mitchell at Figs. 2A and 2B, and at column 5, line 15 to column 6, line 3.

<sup>2</sup> Mitchell at Fig. 2B.

<sup>3</sup> Mitchell at column 6, lines 3-5.

involves sorting to determine the N-best text unit concatenations may not be performed when the token is formed.

Accordingly, Applicants respectfully submit that Mitchell fails to teach or suggest “the processor is further configured to delay a merging of the text unit concatenations in the new token to determine the N-best likelihood values until an end of the speech segment,” as recited in Claim 1, and as similarly recited in independent Claims 13-15, 26, and 27.

In addition, Claim 1 is amended to explicitly recite that the new token includes a pointer to a word history structure that includes pointers to text unit concatenations from previous tokens of different nodes, and offsets indicating a difference in maximum likelihood scores between the text unit concatenations from the previous tokens.

As explained in the background discussion of Applicants’ specification with regard to Figure 9, a conventional system may include two tokens for states 2 and 3, which are passed to a new token for state 5. Further, the token for state 5 may have text unit concatenation entries 1, 2, and 3 which are derived from tokens in states 2 and 3 using a merge operation where the top three entries (e.g., the N-best entries for state 5) are derived by calculating and sorting the text entries for state 2 and state 3. The system in the background discussion is similar to the type of system described by Mitchell in which new tokens, when formed, carry the N text concatenations with the best likelihood values.

However, the conventional system and the system from Mitchell are different than the system of the independent claims. For example, in a non-limiting embodiment of Applicants’ Figure 10, a token in state 5 just contains pointers to the text unit concatenations of the previous tokens and also an offset indicating the difference in the maximum likelihood scores between the text concatenations of the previous tokens. For example, in Figure 10, the top likelihood in state 2 is 150 and the top likelihood in state 3 is 140. Therefore, an offset of -10 is stored. However, beyond this, there is no requirement in the claimed invention for a

sorting process to determine the N-best candidates, and such a sorting process is delayed for as long as possible. Furthermore, Applicants have discovered that by delaying the merging for as long as possible, it becomes advantageously possible for tokens to be pruned and therefore less computer power is needed to sort the list.

As discussed above, Mitchell describes a conventional decoding in which, when two tokens come together, a hypothesis from both of the tokens is combined and sorted so that only the best few are retained. For example, according to a conventional system or the system of Mitchell, if a token can only store a maximum of five hypotheses, then two tokens when combined according to the conventional method or the method of Mitchell can result in more than five hypotheses. Therefore, according to the conventional method or the method of Mitchell, it becomes necessary to sort these more than five hypotheses so that just the best five are retained and carried forward and eventually used to form an N-best lattice. Such a merging process according to the conventional system or the system of Mitchell involves per frame sorting and is therefore costly in terms of computational power. Thus, an embodiment according to the claimed invention may advantageously avoid doing this merging process until it is absolutely necessary, and since some tokens may be pruned away before there is a need to start sorting them in order to determine the N-best lattice, a saving in computational power may be obtained according to the present invention.

Accordingly, Applicants respectfully submit that Mitchell also fails to teach or suggest that “the plurality of text unit concatenations in a token are the text unit concatenations with the N-best likelihood values,” and “said new token including a pointer to a word history structure that includes pointers to text unit concatenations from previous tokens of different nodes, and offsets indicating a difference in maximum likelihood scores between the text unit concatenations from the previous tokens,” as recited in independent Claim 1, and as similarly recited in independent Claims 13-15, 26, and 27.

Accordingly, Applicants respectfully submit that independent Claims 1, 13-15, 26, and 27, and claims depending therefrom, patentably define over Mitchell.

In addition, Applicants respectfully traverse the rejection of Claims 10 and 24 under 35 U.S.C. § 103(a) as unpatentable over Mitchell and Chang.

Claims 10 and 24 depend from independent Claims 1 and 15, respectively, which are believed to patentably define over Mitchell as discussed above. In addition, Applicants respectfully submit that Chang fails to supply the claimed features lacking in the disclosure of Mitchell. Therefore, it is respectfully requested the rejection of Claims 10 and 24 under 35 U.S.C. § 103(a) also be withdrawn.

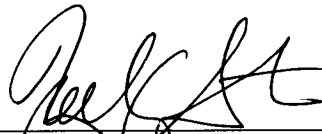
Accordingly, Applicants respectfully submit that independent Claims 1, 13-15, 26, and 27, and claims depending therefrom, are allowable.

Furthermore, it is respectfully noted that the outstanding Office Action, at the cover page, pre-numbered line 10, fails to indicate that the drawings are accepted. Applicants respectfully request a formal indication of acceptance of the drawings in the next Office Action.

Consequently, in light of the above discussion and in view of the present amendment this application is believed to be in condition for allowance and an early and favorable action to that effect is respectfully requested.

Respectfully submitted,

OBLON, SPIVAK, McCLELLAND,  
MAIER & NEUSTADT, P.C.



---

Eckhard H. Kuesters  
Attorney of Record  
Registration No. 28,870

Zachary S. Stern  
Registration No. 54,719

Customer Number

**22850**

Tel: (703) 413-3000  
Fax: (703) 413 -2220  
(OSMMN 08/07)